Institute for Advanced Development Studies

inesad

06/2011

Rethinking Earnings Determinants in the Urban Areas of Bolivia

by:

Beatriz Muriel H.

Development Research Working Paper Series

06/2011

December 2011

The views expressed in the Development Research Working Paper Series are those of the authors and do not necessarily reflect those of the Institute for Advanced Development Studies. Copyrights belong to the authors. Papers may be downloaded for personal use only.

Institute for Advanced Development Studies

inesad

Rethinking Earnings Determinants in the Urban Areas of Bolivia

by

***Beatriz Muriel H.*[‡]**

(December 2010)

Abstract

This paper analyzes the factors that explain earnings in levels and inequality in the urban areas of Bolivia, considering not only the usual individual characteristics (education, experience, gender, and ethnicity) but also firm characteristics. Given the information available at the firm level in the household surveys, two simple models were developed: one for independent workers (for which there is relatively detailed firm-level data), and the other for dependent workers (where firm variables were approximated by sector, size, and by the legal condition of the workers). The main econometric results show that: i) earnings regressions that include only individual variables present highly biased (overestimated) coefficients; ii) firm characteristics are fundamental factors for explaining earnings for independent workers, almost doubling R^2 and explaining 45.5% of observed earnings inequality; and iii) firm proxies for dependent workers are also relevant; however, they explain wages at a lower percentage (11.8%), which may be due to non-detailed firm data and to the relatively higher importance of education for these workers. These new findings represent a contribution to the empirical literature on earnings determinants for urban Bolivia as well as to the vision of labor income and poverty problems.

Keywords: earnings (wages), firm characteristics, inequality, Bolivia

JEL Classification: J20, J23, J31

[‡] Ph.D in economics and senior researcher at INESAD, bmuriel@inesad.edu.bo.

I. Introduction

Several empirical studies have been developed to analyze the factors that explain urban earnings in Bolivia. Generally, the analyses employ variables related to the theory of human capital, such as years of schooling and experience, and controls by gender and ethnic wage gaps (Pérez de Rada 1997, Fields *et al*. 1998, Rivero and Jiménez 1999, Moensted 2000, Andersen 2001, Andersen and Muriel 2002, Spatz and Steiner 2002, Ramírez 2003, Mercado et al. 2003, Muriel 2005a, and Escalante 2004). In many cases, education is pointed out as the most important factor, since the rate of return by year is estimated around 9%, accounting for most of the explained wage inequality.

Abowd *et al.* (1999) and Abowd and Kramarz (1999), however, highlight the importance of considering also the heterogeneous components of firm characteristics in empirical analyses of wages, and demonstrate that the exclusion of these factors could lead to biases in the parameters estimated, due to the omission of variables and/or aggregation problems (use of different linear combinations). The inclusion of these variables empirically can be justified by relaxing an implicit hypothesis of the human capital theory, which is that the labor market is perfect.

The main restriction to an adequate study of factors that explain earnings, however, rests on the need of employer-employee information at the individual level, which in many cases, as in that of urban Bolivia, is not available. In this regard, in this paper a new simple methodology is proposed, adapted to the information presented in the household surveys. This methodology relaxes the hypothesis of a perfect labor market and merges labor personal factors, derived from the human capital theory, and firm factors, derived from the firms' optimization problem.

In this regard, two simple models were developed to be analyzed empirically. The first one is for independent workers – representing 43.1% of the total workers studied – for which there is information on production costs and other firm characteristics (besides the usual individual variables), permitting a detailed evaluation on earnings determinants. The second model is for dependent workers, where firm characteristics are approximated by sector and size as well as by the legal condition of the workers. In particular, this paper highlights the relevance of the wage gaps by sector and size of the firms, even taking into account years of schooling, which supports the alternative hypothesis of imperfections in the labor market.

Earnings have been estimated in levels and variability considering the decomposition analysis of inequality of Shorrocks (1982) and Fields (2002). The empirical analysis confronted three econometric problems: multi-collinearity, endogeneity and sample selection bias. The multi-collinearity arises for the model of independent workers because firm characteristics were

derived from a *translog* production function, which has various variables related to the factors of production. In this regard, the problem was corrected by using Bierrens' (2007) procedure, by evaluating the robustness of the explanatory variables, and finally by including only the factors of production variables that have a low correlation between them and have the highest *t*-values. The endogeneity problem arises because according to the human capital theory, workers' ability is an omitted variable that is correlated to years of schooling. Consequently, instrumental variables were used, estimating the models by Two-Stage Least Squares (2SLS). Lastly, the sample selection bias arises because the division in the employed population between independent workers and dependent ones implies the use of non-randomly selected samples. To correct this bias simultaneously with the endogeneity problem, the Wooldridge (2002) approach was used, which combines Heckman's two-step procedure and instrumental variables.

The econometric results obtained in this paper, by including firm characteristics, significantly contribute to the empirical literature on earnings determinants for urban Bolivia, as well as to the existing vision of low wages, inequality and poverty problems. First, firm characteristics are exposed as fundamental factors for explaining earnings for independent workers: the coefficient of determination (R^2) almost doubles when these variables are considered in the empirical analysis. Additionally, the estimated coefficients of the usual individual variables are highly biased (overestimated) when firm characteristics are excluded, confirming the reflection of Abowd *et al.* (1999) and Abowd and Kramarz (1999). In particular, the results show that independent workers with up to 12 years of schooling have a rate of return of 3.9% per year of education, and those with over 12 years of schooling, a rate of 4.2%. Secondly, firms' proxies for dependent workers are also relevant for explaining wages, and, in the same way as for independent workers, the regressions that only consider individual factors present overestimated coefficients.

Finally, the Sorrocks-Fields inequality decompositions for independent workers show that firm characteristics jointly explain 45.5% of earnings inequality (71.2% considering the Fields decomposition, which excludes the residuals in the calculation), and years of schooling, 6.6%. However, for dependent workers, education is still the most relevant factor explaining earnings dispersion, with a contribution of 28.35%, followed by firms' proxies variables, that jointly explain 11.8%. This last result can be attributed to both non-detailed available information at the firm level and to the relatively higher importance of education for these workers.

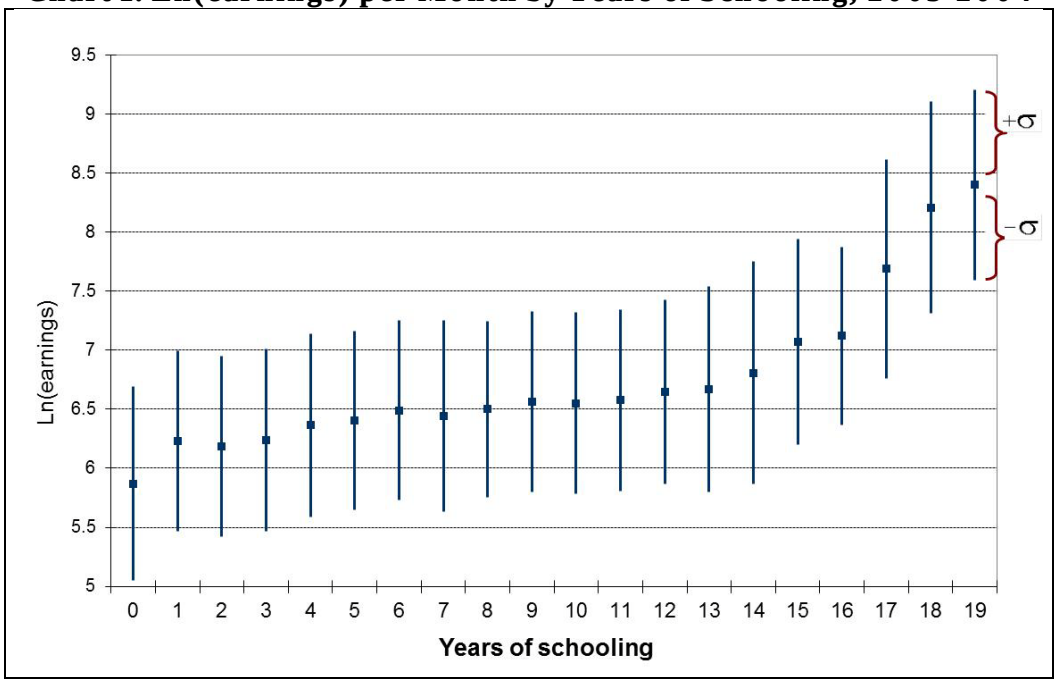
The paper is divided into five sections, including this introduction. The second section shows stylized facts regarding earnings and their interrelation with years of schooling, sectors of work, and size of the firms. The third section describes the methodology for both independent

and dependent workers. The fourth section describes the data, discusses the econometrical problems, and presents the empirical estimations. Finally, the last section describes the most important conclusions.

II. Stylized Facts

The most important factor of incidence on earnings in urban Bolivia, pointed out by the empirical literature, has been education. The returns on years of schooling has been estimated around 9% per year (Pérez de Rada 1997, Fields *et al.* 1998, Rivero and Jiménez 1999, Andersen 2001, Andersen and Muriel 2002, Spatz and Steiner 2002, Ramírez 2003, Mercado *et al.* 2003, and Escalante 2004), and usually explains 50% or more of the explained wages inequality (Fields *et al.* 1998, Andersen 2001, and Spatz and Steiner 2002).

Chart 1: Ln(earnings) per Month by Years of Schooling, 2003-2004



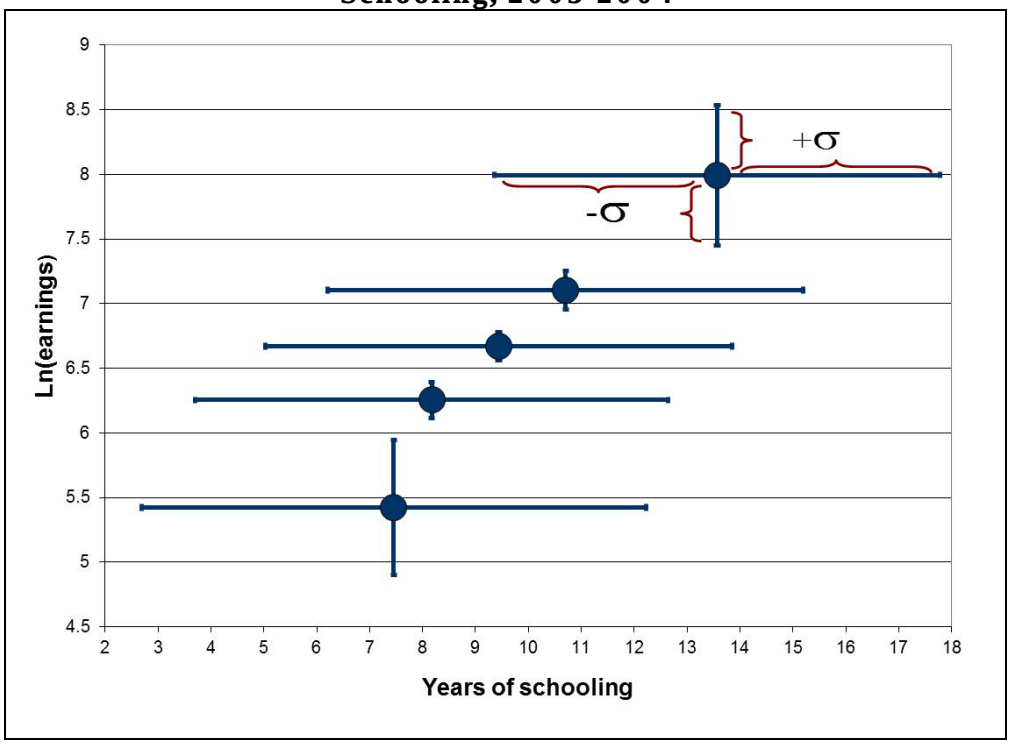
Source: Prepared by author based on Household Survey, 2003-2004.
Notes: 1) The arrows indicate one positive and negative standard deviation with respect to the average. 2) Earnings were adjusted to real values, using the Consumer Price Index, October-November 2004, because the survey was conducted in November 2003 and November 2004. 3) The employed population is between 18 and 65 years of age.

Indeed, Chart 1 shows the positive relationship between the two variables; however, the tendency is non-linear, presenting a higher slope for workers with over 12 years of education (usually having completed high school). This pattern has been previously documented in a very

few studies and controlled in Mincer's-type regressions (Moensted 2000, and Muriel 2005a)[1]. Worth noting is the high variability of earnings by level of education; for example, an individual with a labor income around Bs. 665 per month (ln(665)=6.5) may have zero or up to 16 years of education.

Chart 2 shows a high dispersion in years of schooling when taking the average earnings by quintile. In particular, data suggests that two individuals with an educational gap of 5 years – the first one with 7 and the second with 12 – may earn wages that place both in the first quintile (lowest return) or up to the fourth quintile. Additionally, an employee with 10 or 11 years of schooling may belong to any earning level.

Chart 2: Quintiles of Ln(earnings) per Month by Years of Schooling, 2003-2004



Source: Prepared by author based on Household Survey, 2003-2004.
Notes: 1) The arrows indicate one positive and negative standard deviation with respect to the average. 2) Earnings were adjusted to real values, using the Consumer Price Index of October-November 2004, because the survey was conducted in November 2003 and November 2004. 3) The employed population is between 18 and 65 years of age.

[1] The Household Survey 2003-2004 was chosen because it covered all the months of a year (November 2003-November 2004) and not only one month (November-December) as in the case of household surveys of other years. However, Annex Tables A.1 and A.2, present the econometric results using the survey for 2007, which is the latest year of official information.

Other variables have been included in Mincer's-type regressions to explain the dispersions in the relation between earnings and education, the most important of these (besides experience) being gender and ethnicity. On the one hand, men earn more than women because of: i) occupational segregation problems, where female jobs usually have lower returns; ii) labor cost differences, with pregnancy and postpartum benefits given to women workers by their employers; iii) still a sharp division of responsibilities at home, where men have to generate income to support the family and women have the housework responsibilities, thus working less hours and in less hierarchical positions in the labor market; iv) women with less work experience as compared to men, not properly accounted for the usual experience variable; and v) apparent discrimination problems (Ramírez 2003, and Muriel 2005a). On the other hand, the wage gap by ethnicity, disfavoring indigenous people, can be explained by: i) differences in educational quality against indigenous people; ii) segmentation problems by sector, where indigenous workers are located in low-paid sectors; and iii) possible discrimination problems (Rivero and Jiménez 1999, Andersen and Muriel 2002, and Mercado *et al.* 2003).

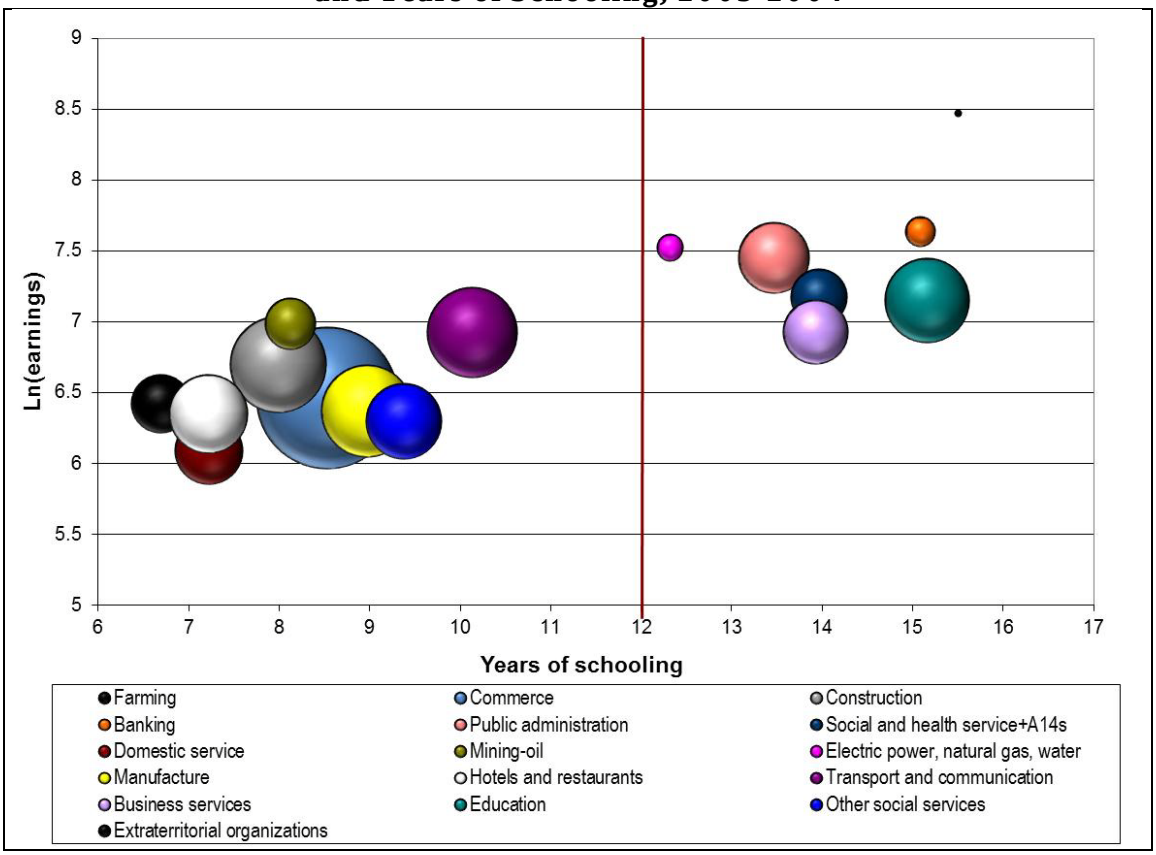
Besides individual characteristics, there are no studies evaluating firm characteristics as determinants of earnings for urban Bolivia[2]. However, two variables at the firm level can be analyzed: sector and size of the firm, given the information available in the household surveys for all the employed population. It is worth noting that these variables seem to be important for explaining earnings. On the one hand, Muriel and Jemio (2010) describe relevant wages gaps by sector, which, in turn, are directly related to labor productivity differences in the way predicted by the theory. On the other hand, Muriel (2010) shows that for the Bolivian manufacturing sector, wages are positively related to labor productivity, and both variables increase as the size of the firm does.

Chart 3 shows average earnings by sector and years of schooling. On average, earnings and years of schooling still have a positive tendency, with sector concentration in both the intensive skilled labor sectors[3] (over 12 years of education) and in the intensive unskilled ones[4] (with up to high school completed). However, some sectors present different average earnings with similar levels of education and others, similar average earnings with different levels of education. Workers in transportation-communications and mining-oil sectors, for example, enjoy earnings similar to those belonging to business services, but with lower schooling levels: 10.2 and 8.1 years respectively, compared to the latter, of 13.9 years. Average labor income in

[2] However, in some cases sector dummies were considered.
[3] Includes the sectors of electricity, natural gas and water supply, public administration, business services, social and health services, banking, education, and extraterritorial organizations.
[4] Includes the sectors of farming, hotels and restaurants, domestic service, construction, mining and hydrocarbons, commerce, manufacturing, other social services, and transport and communication.

farming, manufacturing, and commerce sectors are also similar, even though the education level in the farming sector is lower. Additionally, the electricity, natural gas and water supply, and public administration sectors have higher returns compared with business services, although the latter shows more years of education. Finally, average wages between banking and education sectors are different but with similar levels of education.

Chart 3: Ln(earnings) per Month by Sector and Years of Schooling, 2003-2004



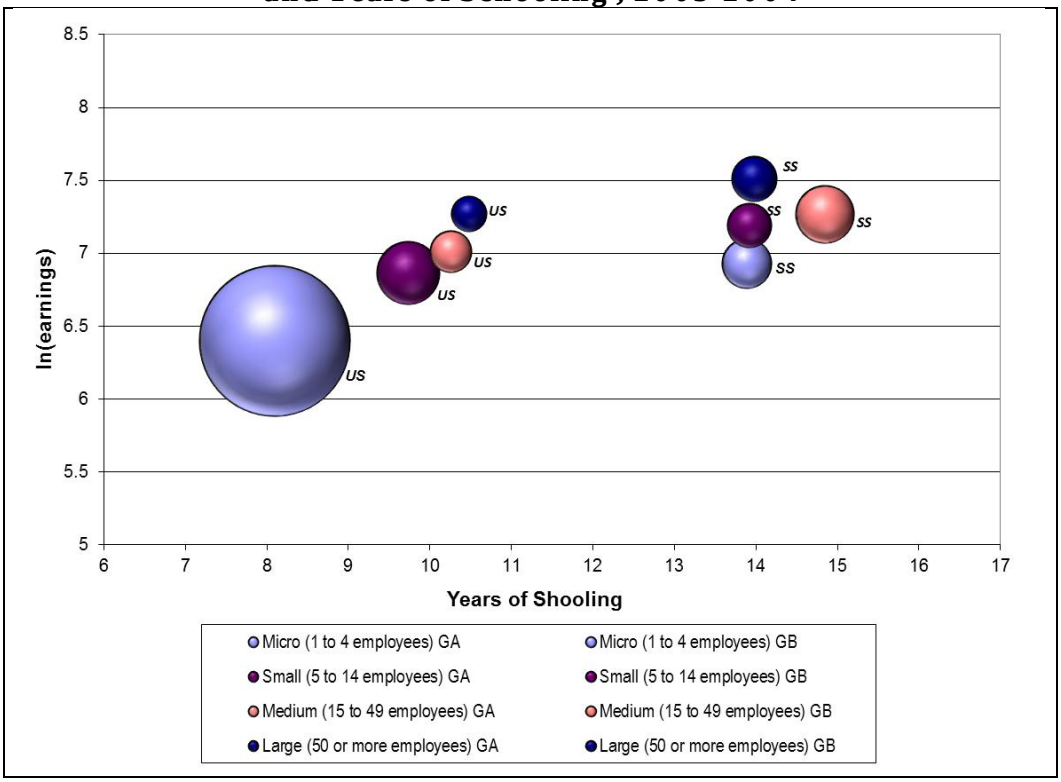
Source: Prepared by author based on Household Survey, 2003-2004.
Notes: 1) The bubble size corresponds to the observations size. 2) Earnings were adjusted to real values, using the Consumer Price Index of October-November 2004, because the survey was conducted during November 2003 and November 2004. 3) The employed population is between 18 and 65 years of age.

Lastly, Chart 4 shows average earnings of workers by firm size and years of education considering two sector groups following the previous Chart: the first one, called *SS*, includes intensive skill sectors; and the second one, *US*, gathers the sectors with the remaining employed population (unskilled). In each group, earnings increase with the size of the firms, however, the relation with education is less clear. Workers of small and medium firms belonging to the *US* group, for example, have an average wage relatively similar than those of micro firms belonging to *SS*, although the former have less years of education –10.3 compared

to 13.9. Additionally, the average wage of workers of large firms of *US* is higher than of micro firms of *SS*, although in the latter case the level of education is higher. Finally, in the *SS* group, workers of large firms have the highest average salary, although the levels of education are lower when compared to those of medium firms and statistically similar (according to the Wald test) to those of micro and small firms.

In short, the information above shows differences (or similarities) between average wages by sector and firm size that are not accurately derived by differences (or similarities) in years of schooling. The next section discusses a way to include firm characteristics as additional earnings determinants, adjusted to the information available for the urban Bolivia case.

Chart 4: Ln(earnings) per Month by Size of Firm and Years of Schooling , 2003-2004



Source: Prepared by author based on Household Survey, 2003-2004.
Notes: 1) The bubble size corresponds to the observations size. 2) Earnings were adjusted to real values, using the Consumer Price Index, October-November 2004, because the survey was conducted in November 2003 and November 2004. 3) The employed population considered is in between 18 and 65 years of age. 4) GA corresponds to sectors with average years of education over 12 according to Chart 3, and GB is the remaining sectors.

III. Modeling Earnings Determinants

III.1. Empirical Issues

Abowd *et al.* (1999) and Abowd and Kramarz (1999) highlight the importance of taking into account the heterogeneous components of firm characteristics in empirical analyses of wages (besides usual individual characteristics related to the theory of human capital), and demonstrate that the exclusion of these factors could lead to high biases in the parameters estimated, which are the result of omissions of variables and/or aggregation problems (use of different linear combinations).

The main restriction to an adequate study of the factors that explain earnings, however, rests on the need for information at worker level of both individual characteristics and firm characteristics. This data has been available in a few cases, usually for developed countries or for specific studies (see, for example, Verner 1999, Jones 2001, Soderbom *et al.* 2005, Martins 2008, and Aydemir and Skuterud 2008). In other cases, empirical literature has been creative to match the two variable groups at the worker level in order to study different relevant topics concerning earnings. Some labor surveys that have firm identifiers have been used to match with the corresponding establishment surveys (see, for example, Abowd and Kramarz 1997, Abowd *et al.* 1999, Hellerstein and Neumark 2004, Lallemand *et al.* 2005, Ilmakunnas and Maliranta 2005). Administrative records of workers formally employed have been also matched with the corresponding firms' information, which in turn is derived from either establishment surveys or tax statements (see, for example, Haltiwanger *et al* 2000, Menezes-Filho *et al.* 2006, and Woodcock 2007 and 2008). Some labor surveys have information on the type of business or industry and work location of people, which allowed assigning an industry-location cell for every worker and matching it with firms that also have this information (Troske 1995, Hellerstein *et al.* 1999, Moretti 2004, and Breau and Rigby 2006). Finally, available firm characteristics, such as sector, size or work area, have been included in Mincer's type regressions (see, for example, Brown and Medoff 1989, Oi and Idson 1999, Bayard and Troske 1999, Herrera 2003, Muriel 2005a, Temesgen 2005, Lallemand *et al.* 2005, and Muravyev 2009). [5]

In the case of urban Bolivia, the household surveys have ample information on labor at the individual level and include some characteristics of firms, such as size and sector for all workers and production costs for independent workers. Establishment surveys, in contrast, are scarce for the manufacturing sector and almost inexistent for other industries. One advantage, however, is that independent workers (that have production cost information) represent a

[5] See also Abowd and Kramarz (1999) and Abowd *et al.* (2008) for surveys.

considerable portion of the employed population analyzed, 43.1% for 2003-2004. In this context, a methodology is proposed, presented in the next subsection, considering independent workers and dependent workers separately, with specific estimation equations for each case.

III.2. Methodology

The empirical relationship between earnings and years of schooling is derived from the theory of human capital (Becker 1964), which assumes implicitly that the labor market is perfect: two individuals with the same education should have the same wage, and the same marginal productivity value as a result of labor market equilibrium, regardless of dissimilar firm characteristics. Differences in education are then the source for explaining differences in earnings and productivity. The simplest form of the model exposes these results (see Mincer 1974): the wage gap between choosing s or $\bar{s}$ years of education is given by

$$(1) \qquad w_s = w_{\bar{s}} e^{r(s-\bar{s})}$$

where $r>0$ is the discount rate, and w_s ($w_{\bar{s}}$) is the wage for workers with s ($\bar{s}$) years of schooling. Note that if $s > \bar{s}$ then $w_s > w_{\bar{s}}$ and if $s = \bar{s}$ then $w_s = w_{\bar{s}}$. Expression (1) has been refined supposing that marginal returns and costs (or tastes) to schooling are heterogeneous between workers, modeling abilities and including other relevant individual characteristics (see Mincer 1974 and Card 1999). Nevertheless, for simplicity's sake without loss of generality, it is assumed initially that workers are different only by the years of schooling that they choose (this assumption is later relaxed).

When the labor market is not perfect, however, (1) does not hold because wages can be different even for workers with the same education or equal for workers with different years of schooling. In this regard, (1) can be redefined in a simple way by supposing what follows. First, firms are gathered in $\overline{K}$ groups corresponding to the segmentation groups of the labor market, where two groups of firms K and K' pay different wages for the same level of education, and two firms, k and k', within a group K pay the same wage for the same level of education. Second, workers do not know *a priori* in which group of firms they will work and they form in the same way their salary expectations (because in the simple form of the human capital theory people finish their formal schooling before entering the labor market). Therefore (1) is redefined as:

$$(1') \qquad E(w_s) = E(w_{\bar{s}}) e^{r(s-\bar{s})}$$

where $E(w_s)=\sum_{K=1}^{\overline{K}}\pi_K w_{s,K} \quad \forall s$, $\sum_{K=1}^{\overline{K}}\pi_K=1$, $w_{s,K}$ and π_K are respectively the salary and the probability of working in a firm k (=1,2, ..., k_K) belonging to group K(= 1, 2,..., $\overline{K}$).[6] It is worth noting that for any firm k belonging to group K the following equation applies:

(1") $\quad w_{s,K}=w_{\bar{s},K}e^{r(s-\bar{s})}=w_{\bar{s},K}g_s(\cdot)$

where $g_s(\cdot)=e^{r(s-\bar{s})}$. Note that under the assumptions made, the human capital theory is valid for firms within each group K. The salary $w_{s,K}$ correspond to the workers with s years of education and, as will be seen below, $w_{\bar{s},K}$ can be approximated to the average wage corresponding to the average of years of schooling of workers in firm k belonging to K, which, in turn, equals the average labor marginal productivity value[7].

The average labor marginal productivity value is derived considering generic technology of production $F(\boldsymbol{x}_k;\boldsymbol{d}_k)$ for firm k, where $\boldsymbol{x}_k$ represents the vector row of inputs and $\boldsymbol{d}_k$ the vector row of other relevant characteristics of the firms. The maximization problem for firm k implies that:

(2) $\quad w_{\bar{s},K}=\varphi_k p_k\frac{\partial F(\boldsymbol{x}_k;\boldsymbol{d}_k)}{\partial x_{l,k}}\cong\varphi_k\frac{p_k y_k}{\partial x_{l,k}}\frac{\partial \ln y_k}{\partial \ln x_{l,k}}\cong\varphi_k\frac{p_k y_k}{x_{l,k}}\varepsilon_{l,k}$

where $x_{l,k}$ represents employment, φ_k is the degree of market power, p_k the level of prices, y_k the production volume, and $\varepsilon_{l,k}$ the employment-product elasticity. Expression (2) can be substituted into (1") to find the wage at the individual level, which is established as a proposition.

Proposition: *The wage of the individual i (=1, 2, .., i_k) with s years of education, that works in a firm k from grou*

p K, is determined as:

---

[6] Note that probabilities can change depending on the years of schooling; for example, skilled workers can have a greater possibility of working in skill-intensive sectors (if the groups are divided by sectors). However, it is not relevant for the methodology proposed because it analyzes wages once workers are in the labor market.

[7] The literature describes many reasons for imperfections in the labor market (such as unions, minimum wages applied only to the formal sector, efficient salaries, implicit contracts, etc.) that lead to different wage levels in equilibrium. In order to avoid this discussion that leads away from the objective of the paper, it is assumed here that wages are always within the labor demand curve, which means that wages equal the value of labor marginal productivity.

(3) $w_{is,K} = w_{\bar{s},K}\, g_{is}(\cdot)$

where $g_{is}(\cdot) = e^{r(s-\bar{s})}$ *values the education gap between worker i and employment in the firm (on average) as a result of the marginal productivity gap between them.*

Proof: Employment can be redefined as: $x_{l,k} = \sum_i x_{il,k} = \sum_i \tilde{g}_{is} x_{il,k}$ (condition 1), where $x_{il,k}$ is the labor supply given by worker *i* (in units or hours) and $\tilde{g}_{is}$ are weights that allow adding workers in "productivity equivalent units"[8] and, at the same time, reflect different years of schooling between workers. To illustrate this point, it is assumed that two workers, *i* and *i*', offer one unit of labor. If the education levels are higher for the former (*s*) than for the latter (*s*'), then in "productivity equivalent units" worker *i* represents more labor units than worker *i*' and $\tilde{g}_{is} > \tilde{g}_{i's'}$ (condition 2); for example one unit of labor of worker *i* can represent double that of worker *i*', which implies: $\tilde{g}_{is} = 2\tilde{g}_{i's'}$. In this regard, the wage for worker *i* is derived from the maximization problem[9]:

(3') $$w_{is,K} = \varphi_k p_k \frac{\partial F(\boldsymbol{x}_k; \boldsymbol{d}_k)}{\partial x_{l,k}} \frac{dx_{l,k}}{dx_{il,k}} \cong \varphi_k \frac{p_k y_k}{x_{l,k}} \varepsilon_{l,k} \tilde{g}_{is} \cong w_{\bar{s},K} \tilde{g}_{is} = w_{\bar{s},K}\, g_{is}(\cdot)$$

where consequently the following are used: expression (2), condition 1 and expression (1"), from which it follows that weights $\tilde{g}_{is} = g_{is}(\cdot) = e^{r(s-\bar{s})}$, meeting conditions 1 and 2. □

As mentioned above, from equation (3) two empirical models are constructed according to the available data for the urban Bolivia case. The first one corresponds to independent workers; that is, self-employed persons, managers or partners, and members of production cooperatives that are not salaried employees, and who have information on their own production units. In these cases it is possible to have a detailed analysis of variables that explain $w_{is,K}$. The second model corresponds to dependent workers who do not have information at the firm level, but do have information on the sector and size of the firms where they work.

For the first empirical model, equation (3) is redefined considering that the production function is a *translog*, assuming that crossed products between inputs and own characteristics do not exist, and applying logarithms:

(3') $$\ln w_{is,K} = \gamma_{0,k} + \gamma_1 \ln \varepsilon_{l,k} + (\gamma_l - 1) \ln x_{il,k} + \sum_{j \neq l} \gamma_j \ln x_{j,k}$$

$$+ \frac{1}{2} \sum_j \sum_{j'} \gamma_{jj'} \ln x_{j,k} \ln x_{j',k} + \sum_d \gamma_{d,} d_{d,k} + \ln g_{is}(\cdot)$$

---

[8] This approximation follows Trefler (1993).

[9] The derivative between total employment and employment *i* would be applicable when labor supply is measured in hours, workers *i* (in units) represents a group of workers with the same level of education, or the firm hires an infinitesimal unit of labor which has the labor performance of *i*.

where sub-indexes *j*, *j*' are attributed to inputs; $\gamma_{0,k} = \ln \varphi_k + \gamma_0$; $\gamma_{jj'} = \gamma_{j'j}\ \forall$ *j*, *j*' and *j*≠ *j*' $\sum_j \gamma_{jj'} = 0$; $\sum_{j'} \gamma_{j'j} = 0$; and $\sum_j \frac{\gamma_j}{\varpi} = 1 - \varpi$ represents the degree of homogeneity of the production function (Berndt and Christensen 1973, and Binswanger 1974).

Since data of firm characteristics is collected at the level of independent workers (one unit of employment), it is considered that these represent a kind of employment different from the rest, which means separating this employment ($x_{il,k} \approx 1$) from the rest of workers, and redefining expression (3') as:

$$(3'') \qquad \ln w_{is,K}\Big|_{x_{il,k}=1} = \gamma_{0,K} + \gamma_1 \ln \varepsilon_{l,k} + \sum_{j \neq il} \gamma_j \ln x_{j,k} + \frac{1}{2} \sum_{j \neq il} \sum_{j' \neq il} \gamma_{jj'} \ln x_{j,k} \ln x_{j',k}$$

$$+ \sum_d \gamma_{d,} d_{d,k} + \ln g_{is}(\cdot)$$

where it is assumed that $\gamma_{0,k} = \gamma_{0,K}$, $\forall$ for all firms *k* that belong to group *K*, which means that market power is homogeneous among firms belonging to a certain group.

Function $\ln g_{is}(\cdot)$ is the well-known Mincer's-type regression; that is, the part of the wage that is explained by observable individual factors. Now the human capital theory (equation (1)) is relaxed, allowing to add other individual relevant characteristics and to control for non-linearity in education returns:

$$(4) \qquad \ln g_{is}(\cdot) = \alpha_0 + \alpha_1 s_{1i} + \alpha_2 s_{2i} + \alpha_3 e_i + \alpha_4 e_i^2 + \sum_n \xi_n d_{i,n}$$

where s_{1i} and s_{2i} are respectively the measures of years of schooling up to 12 years (zero otherwise) and over 12 years; e_i and e_i^2 represent experience and squared experience of worker *i*; $d_{i,n}$ is the n^{th} relevant attribute of worker *i*; and the alphas and xs are the coefficients. The first empirical model is then established by substituting (4) into (3"):

$$(5) \qquad \ln w_{is,K}\Big|_{x_{il,k}=1} = \tilde{\gamma}_{0,K} + \gamma_1 \ln \varepsilon_{l,k} + \sum_{j \neq il} \gamma_j \ln x_{j,k} + \frac{1}{2} \sum_{j \neq il} \sum_{j' \neq il} \gamma_{jj'} \ln x_{j,k} \ln x_{j',k}$$

$$+ \sum_d \gamma_{d,} d_{d,k} + \alpha_0 + \alpha_1 s_{1i} + \alpha_2 s_{2i} + \alpha_3 e_i + \alpha_4 e_i^2 + \sum_n \xi_n d_{i,n} + u_i$$

where $\tilde{\gamma}_{0,K} = \gamma_{0,K} + \alpha_0$ and u_i is the error term.

The second empirical model corresponds to dependent workers, where the wage of group *K* is approximated by sector, size and formal condition of the workers (see data description sub-section). The salary of a worker *i* is determined by using (3) and (4):

$$(6) \qquad \ln w_{is,K} = \tilde{\alpha}_0 + \alpha \ln w_{\bar{s},K} + \tilde{\alpha}_1 s_{1i} + \tilde{\alpha}_2 s_{2i} + \tilde{\alpha}_3 e_i + \tilde{\alpha}_4 e_i^2 + \sum_n \tilde{\xi}_n d_{i,n} + \upsilon_i$$

where the alphas and xs are the coefficients to be estimated, and υ_i is the error term.

Lastly, equations (5) and (6) are simplified for further discussion:

$$(7) \qquad \tilde{w}_{is,K} = \sum_m \beta_m \tilde{x}_{im} + \nu_i$$

where $\tilde{w}_{is,K} = \ln w_{is,K}$, β_m (m=0, 1, 2,.., M) is the estimated coefficient corresponding to the explanatory variable $\tilde{x}_{i,m}$, where $\tilde{x}_{i,1}$=1, and ν_i is the error term.

III.3. Inequality Decomposition

A second important issue lies in the decomposition analysis of inequality, which allows evaluating the main determinants of earnings inequality. In this regard, the methodologies of Shorrocks (1982) and Fields (2002) are used (see also Cowell and Fiorio 2009)[10]. From (7), define $\theta_m = \frac{COV(\beta_m \tilde{\boldsymbol{x}}_m, \tilde{\boldsymbol{w}}')}{VAR(\tilde{\boldsymbol{w}})}$ as the "relative factor inequality weight", and $\theta_\nu = \frac{COV(\boldsymbol{\nu}, \tilde{\boldsymbol{w}}')}{VAR(\tilde{\boldsymbol{w}})}$, where $\tilde{\boldsymbol{x}}_m, \tilde{\boldsymbol{w}}, \boldsymbol{\nu}$ are row vectors. Then applying variance to (7) and dividing by $VAR(\tilde{\boldsymbol{w}})$ gives the inequality decomposition normalized to 1 (or 100%):

$$(7') \qquad 1 = \sum_m \theta_m + \theta_\nu$$

where $\theta_m = \beta_m^2 \frac{VAR(\tilde{\boldsymbol{x}}_m)}{VAR(\tilde{\boldsymbol{w}})} + \sum_{m' \neq m} \beta_m \beta_{m'} \frac{COV(\tilde{\boldsymbol{x}}_m, \tilde{\boldsymbol{x}}'_{m'})}{VAR(\tilde{\boldsymbol{w}})} + \beta_m \frac{COV(\tilde{\boldsymbol{x}}_m, \boldsymbol{\nu}')}{VAR(\tilde{\boldsymbol{w}})}$

[10] The methodology has been criticized when the dependent variable is in logarithms, which may break up (possibly in a rare case) the transfer axiom: if a new distribution is obtained from another one by taking income from the less-paid worker to the better-paid one, measured inequality should increase. However, in the paper the method is used because there is not yet a better methodology for regression-based inequality decomposition (for a discussion of the methodologies, see, for example, Morduch and Sicular 2002, Cowell and Fiorio 2006, and Israeli 2007).

and $\theta_{\nu} = \frac{VAR(\boldsymbol{\nu})}{VAR(\tilde{\boldsymbol{w}})} + \sum_{m} \beta_m \frac{COV(\tilde{\boldsymbol{x}}_m, \boldsymbol{\nu}')}{VAR(\tilde{\boldsymbol{w}})}$. Notice that if the exogenous variables are not correlated between them or with the error term, the expressions are reduced respectively to $\theta_m = \beta_m^2 \frac{VAR(\tilde{\boldsymbol{x}}_m)}{VAR(\tilde{\boldsymbol{w}})}$ and $\theta_{\nu} = \frac{VAR(\boldsymbol{\nu})}{VAR(\tilde{\boldsymbol{w}})}$. Additionally, Fields (2002) ignores the error component and re-normalizes (7') where 100% corresponds to the inequality decomposition of the explained dependent variables (corresponding to the coefficient of determination R^2).

IV. Empirical Analysis

IV.1. Econometric Issues

The estimation of the models of earnings determinants presents three potential econometrics problems: multi-collinearity (for the case of independent workers), endogeneity and sample selection bias.

The multi-collinearity problem arises because several explanatory variables derived from the *translog* production function are highly correlated, which tends to inflate the standard deviations (deflating the *t*-values) as well as the coefficient of determination R^2. Bierrens (2007) shows two possible solutions for this problem. The first one is to reduce the number of explanatory variables, by imposing restrictions on the parameters based on the theory. The second solution is to delete the variables from the model that cause the problem. Here, the second option is chosen because the data is not adequate enough to determine the production function structure; for example, the returns of scale (principally the proxy of the capital stock that will be described later). Because the *t*-values cannot be used for analyzing which variables should be deleted, Bierrens proposes applying the following steps: i) to choose the least relevant variable for the analysis (that is also insignificant); ii) to regress this variable over the rest of the explanatory variables; iii) to estimate the residuals (which are not correlated with the rest of the explanatory variables); iv) to replace the explanatory variable chosen in i) by the residuals estimated in the main regression, which re-parameterizes the model; v) if there are some variables that are still insignificant, to repeat the procedure until all the *t*-values of the remaining variables are significant. In addition to Bierrens' methodology, the robustness of the explanatory variables will be analyzed by estimating the model with different sub-groups of explanatory variables.

The endogeneity problem arises because – according to the human capital theory – workers' ability is an omitted variable (unobserved) that is correlated with the years of schooling, which

leads to a biased estimated coefficient for education[11]. In this regard, instrumental variables are used, taking as instruments institutional features of the school system or family background, such as the education of parents or spouses. The household surveys used do not have this kind of data for all the workers analyzed, so that alternatively two instrumental variables for the urban Bolivia case are proposed corresponding to the two variables of education (see equation 4). The first one is the family average years of schooling. The variable is constructed considering the average years of education of the family members of 18 years of age or more (excluding the worker), which are categorized as: household head, wife, son or daughter, son-in-law or daughter-in-law, and parents. 87% of workers have at least one family member (besides themselves) so it was possible to calculate the average years of schooling for them. For the remaining 13%, where the workers are the only adults at home, are other relatives or are domestic workers (not belonging to the family) their corresponding years of schooling was used[12]. This instrument has a justification similar to that of other family variables: education of members of a family are correlated because spouses share mutual interests and behaviors, and parents want all their children to reach at least the same level of education as they. The second one assumes that most of the employed population had restricted access to university (institutional limitations) because: i) universities are located in the main cities; ii) universities do not have entrance examinations; however, they are economically restricted for poor people (the opportunity cost is very high for them); and iii) even some of the employed population that did not completed high school faced location and monetary restrictions which limited their access to university. In this regard, a dummy is constructed, being equal to one if the worker is a post-graduate, has completed tertiary school at university, or is currently attending university, and zero otherwise.

Empirically, two additional equations are added for years of schooling with a similar structure:

$$(8) \qquad s_i = \delta_0 + \delta_1 z_{1i} + \delta_2 z_{2i} + \eta_i$$

where $s_i = s_{1i}, s_{2i}$, z_{1i} and z_{2i} are respectively the family average years of schooling variable and the dummy for university, and η_i is the error term with the usual properties. In this regard, two systems of equations are constructed corresponding to independent and dependent workers and used Two-Stage Least Squares (2SLS).

---

[11] However, where workers have heterogeneous costs and returns, other kinds and causalities of biases arise (see Card, 1999).

[12] Many alternatives were evaluated empirically, such as including the workers in the calculation, and/or excluding the remaining 13% of workers. However these modifications did not changed the main econometric results.

The last apparent econometric problem, sample selection bias, arises because the division of the employed population between independent workers and dependent ones implies the use of non-randomly selected samples conducting to missing data problems (see Heckman, 1979). To correct for this bias, the literature usually uses Heckman's two-step procedure. However, because the models analyzed here have two biases at the same time – endogeneity and sample selection – the Wooldridge (2002) approach is used, which combines Heckman's two-step procedure and instrumental variables[13]. First, a probit model is estimated for independent workers (or dependent ones) from the entire urban working-age population (10 or more years of age according to the Instituto Nacional de Estadística, INE), as a function of all relevant characteristics including the instruments and excluding the years of schooling variables:

$$(9) \qquad P(q_i = 1 | \tilde{\tilde{\tilde{\boldsymbol{x}}}}_i) = \Phi(\boldsymbol{\zeta}' \tilde{\tilde{\tilde{\boldsymbol{x}}}}_i)$$

where q_i is equal to 1 if the individual in the working age population is an independent worker for the model (5) (dependent worker for the model (6)) and is between 18 and 65 years of age and zero otherwise, and $\tilde{\tilde{\tilde{\boldsymbol{x}}}}_i$ is a vector of the explanatory relevant variables. Second, the estimated coefficients are obtained using all observations, which in turn, are used to estimate the inverse Mills ratio $\hat{\lambda}_i = \lambda(\hat{\boldsymbol{\zeta}}' \tilde{\tilde{\tilde{\boldsymbol{x}}}}_i)$ for the sub-sample of the working-age population studied.

Third, (7) is redefined including the inverse Mills ratio:

$$(7") \qquad \tilde{w}_{is,K} = \sum_m \beta_m \tilde{x}_{im} + \beta_\lambda \hat{\lambda}_i + \tilde{\nu}_i$$

Finally, the two equations described by (8) and (7") are estimated jointly using 2SLS. And, statistical tests are estimated to analyze the pertinence of both the endogenous problems, through the Hausman test, and the selection problem, by the statistical significance of $\hat{\beta}_\lambda$. Wooldridge shows that this procedure is consistent; however the *t*-values should still be corrected by the "generated regressors problems".

IV.2. Data Description

The information required to empirically analyze equations (5) and (6) was obtained from the Household Survey, 2003-2004. This database is used because it is the richest in information, since it was conducted throughout an entire year, November 2003-November 2004, as

[13] From a previous use of the methodology; see, for example, Garcia *et al* (2001), Das *et al* (2003), Reza and Mussurov (2006), and Chen and Shigeyuki (2009).

opposed to other household surveys that were conducted only for one month. However, Annex Tables A.1 and A.2 present the estimation of both models for the last available official Household Survey, 2007, with similar econometric results. The variables related to firm characteristics for independent workers (self-employed persons, managers or partners, and members of production cooperatives that are not salaried employees) are:

- The neperian logarithm of monthly earnings[14], which is calculated by the Instituto Nacional de Estadística (INE) and corresponds to the net profit by deducting from the gross income all declared costs (payment of salaries, intermediate consumption, service expenses, tax payments, and other contributions). The monthly earnings dispersion derived from differences in hours worked per month is controlled by including them explicitly in the regressions, *hours_month*[15].

- *Ln(product-labor elasticity)*, which is obtained noting that the elasticity is equal to the share of the labor costs over the total cost (see, for example, Yasar *et al*. 2008), approximated by monthly earnings divided by the gross income.

- The factors of production: *ln(remaining workers)* estimated as the total number of workers in the firm minus one (which corresponds to the independent worker); a proxy variable for *ln(capital)* corresponding to the payments for services such as rent, water, electricity, phone, security, and others related; and *ln(intermediate consumption)* that sums up the purchase of raw materials, tools, equipment, and merchandise.

- Dummy variables. A dummy for credit (*D_credit*), equal to one when the independent worker declares having obtained a loan in cash for her business in the last 12 months. 15 dummies for the 16 sectors: farming, mining-oil, commerce, electricity, natural gas and water supply, construction, manufacturing, hotels and restaurants, transport and communication, banking, business services, public administration, education, social and health services, other social services, domestic service, and extraterritorial organizations. In this case, these variables would control, among other things, the market power related to the sector ($\gamma_{0,K}$). *D_has RUC* for firms registered under the Registro Único de Contribuyentes, which aims to capture the best productive and demand scenarios that motivate independent workers to become formal[16]. A dummy for the main cities,

[14] Wages are determined by month and not by hours, since information on (remaining) employment are in units.
[15] This variable is also considered for the case of dependent workers.
[16] In Bolivia very few firms are registered with a RUC number (currently NIT) because it generally implies payment of several taxes and transaction costs related to public bureaucracy. The advantages of having a RUC are that firms can engage in legal exports as well as operating with large public and private firms.

D_main cities, equal to one for cities within the central axis — La Paz, Cochabamba, and Santa Cruz — and zero otherwise. The variable should be positively related to earnings due to higher prices of goods and services in this region, as well as for its relatively greater dynamics and economic development[17].

In the case of dependent workers, the salaries were calculated as the neperian logarithm of monthly earnings, estimated by the INE, which includes declared salaries, bonuses, Christmas bonuses, primes, etc. The two proxies of firm characteristics are:

- *Ln(average wages by sector-size)*, estimated by dividing the earnings of the employed population (dependent workers and independent ones) by sector of work (16) and size of firm (micro, small, medium, and large). Sizes were accounted for all sectors except for banking, public administration, education, social and health services, domestic service, and extraterritorial organizations[18]. The division gave 46 groups for the entire sample.

- A dummy *D_subject to labor legislation* was constructed as firms' characteristic. This variable is a proxy of workers who work under labor legislation; its objective is to capture benefits from both having "decent work" and enjoying better productive conditions (since formalization of firms can be associated with production characteristics that are better and more profitable). In this regard, this variable is related to higher personal effort and better production structures. The variable was constructed considering workers that both contribute to the AFPs (Pension Funds Administrators) and get Christmas bonus (excluding those from public administration which are unstable jobs in Bolivia), and workers that both do domestic services and receive Christmas bonus (which have an special labor legislation).

The variables of individual characteristics include: i) up to 12 *years of schooling* and *over 12 years of schooling* in order to capture the non-linearity of education observed in Chart 1; ii) *Experience* (equal to the worker's age minus years of schooling minus six) and its square; and iii) dummy variables for male condition, *D_male*, and indigenous condition, *D_indigenous* (measured by the worker native language). Lastly, the probit regressions included also: i) a dummy for household head; ii) a dummy for married; iii) a dummy for student; and iv) a ratio of children (up to 9 years of age) per adult (over 18 years of age) by house applied only to

[17] This variable is also considered for dependent workers.
[18] The banking sector was not divided by size due to few observations, nor were the remaining sectors, since they are associated to public administration and international organizations that enforce salary policies regardless of firm size.

women, which is as a proxy variable of caring for children at home and has a negative effect on female participation in the labor force (see Muriel 2005b).

The variables were constructed for workers ages 18 to 65 that declared a positive labor income (excluding some extreme values). The data was adjusted to real values on the basis of the Consumer Price Index, October-November 2004 since the survey was conducted between November 2003 and November 2004. Summary statistics for independent and dependent workers are given in Tables 1 and 2 respectively.

Table 1: Summary Statistics for Independent Workers (4251 observations, 43.1% of the sample)

Variable	Mean	Standard Deviation	Pearson coefficient of variation
Ln(earnings)	6.477	0.959	0.148
Years of schooling	8.172	4.862	0.595
Proportion: workers with years of schooling up to 12	0.836		
Proportion: workers over 12 year of schooling	0.164		
Family average years of schooling	8.910	4.494	0.504
Proportion: dummy for university equal 1	0.112		
Experience	25.866	13.672	0.529
[Experience]2	855.956	780.417	0.912
Proportion of males	0.502		
Proportion of indigenous	0.378		
Hours_month	194.941	103.354	0.530
Ln(employment-product elasticity)	-0.691	0.639	-0.924
Ln(remaining workers)	0.313	0.943	3.016
Ln(capital)	1.081	2.140	1.980
Ln(intermediate consumption)	4.960	2.990	0.603
Ln(capital)xln(intermediate consumption)	7.061	15.393	2.180
Ln(capital)xln(remaining workers)	0.451	2.038	4.514
Ln(intermediate consumption))xln(remaining workers)	1.609	5.210	3.238
$[Ln(capital)]^2$	5.748	12.826	2.231
$[Ln(intermediate\ consumption)]^2$	33.542	24.845	0.741
$[Ln(remaining\ workers)]^2$	0.986	5.554	5.633
Proportion: dummy has RUC equal 1	0.131		
Proportion: dummy credit equal 1	0.182		
Proportion: dummy main cities equal 1	0.567		
Proportion: dummy household head equal 1	0.589		
Proportion: dummy married equal 1	0.752		
Proportion: dummy student equal 1	0.068		
Ratio of children per adult for women	0.221	0.210	0.952

Table 2: Summary Statistics for Dependent Workers (5608 observations, 56.9% of the sample)

Variable	Mean	Standard Deviation	Pearson coefficient of variation
Ln(earnings)	6.849	0.867	0.127
Years of schooling	11.152	4.647	0.417
Proportion: workers with years of schooling up to 12	0.604		
Proportion: workers over 12 year of schooling	0.396		
Family average years of schooling	9.981	4.607	0.462
Proportion: dummy for university equal 1	0.248		
Experience	16.991	11.960	0.704
[Experience]2	431.715	545.692	1.264
Proportion of males	0.621		
Proportion of indigenous	0.211		
Hours_month	204.616	86.726	0.424
Ln(average wages by sector-size)	7.214	0.473	0.066
Proportion: dummy subject to labor legislation equal 1	0.326		
Proportion: dummy main cities equal 1	0.577		
Proportion: dummy household head equal 1	0.538		
Proportion: dummy married equal 1	0.625		
Proportion: dummy student equal 1	0.167		
Ratio of children per adult for women	0.071	0.149	2.117

IV.3. Econometric results

Table 3 shows the results of the empirical analysis on earnings determinants for independent workers. The first three regressions include only the observable individual characteristics, which are related to the theory of human capital: regression (1) is the OLS estimation, (2) corrects for endogeneity bias, and (3) corrects for endogeneity and sample selection biases at the same time. The coefficients estimated show the expected signs, in all cases being consistent with those preciously found in the literature: *years of schooling* have a positive effect on earnings with a rate of return higher for workers with higher years of education – over 12, and *experience* also presents a positive relationship with decreasing marginal returns. Additionally, gender and ethnicity dummies show that males earn more than females and that the indigenous population earns less than the non-indigenous one. It is worth noting that both endogeneity and sample selection biases are relevant in this case: the former was determined

by the Hausman test and the later evaluating the significance of the coefficient of the inverse Mills ratio[19].

Regressions (4) to (6) add to the analysis of the firm characteristics. Initially the multi-collinearity problem was evaluated. In this regard, using the Bierrens (2007) procedure, only one variable was excluded: *ln(capital)×ln(remaining workers)*. However, because of the high correlation between some factor of production variables, and in order not to overestimate the firm characteristics, nor the coefficient of determination used for the decomposition analysis, only the factor of production variables that have a low correlation between them and have the highest *t-values* in the regressions were included. Additionally, the reaming variables included were those that were significant up to a level of 10%.

Table 3: Ln(earnings) Determinants per Month for Independent Workers, 2003-2004

Explanatory variables	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV (2SLS)	IV plus Heckman	OLS	IV (2SLS)	IV plus Heckman
Years of schooling, up to 12	0.042	0.088	0.078	0.019	0.044	0.038
	(9.54)	(7.58)	(6.67)	(5.90)	(4.89)	(4.24)
Over 12 year of schooling	0.058	0.091	0.090	0.024	0.041	0.041
	(15.53)	(13.60)	(13.38)	(7.81)	(7.12)	(7.07)
Experience	0.032	0.036	0.025	0.019	0.0208	0.0140
	(8.93)	(9.89)	(6.39)	(6.98)	(7.46)	(4.79)
[*Experience*]^2	-0.0005	-0.0004	-0.0003	-0.0003	-0.0003	-0.0002
	(-7.85)	(-6.76)	(-4.46)	(-6.11)	(-5.28)	(-3.47)
D_male	0.480	0.409	0.364	0.275	0.242	0.2155
	(19.04)	(13.95)	(12.23)	(12.18)	(9.87)	(8.71)
D_indigenous	-0.198	-0.083	-0.172	-0.125	-0.073	-0.127
	(-7.32)	(-2.41)	(-4.65)	(-6.41)	(-2.94)	(-4.77)
Hours_month	0.003	0.003	0.003	0.001	0.002	0.001
	(4.87)	(4.37)	(-0.37)	(15.23)	(15.35)	(15.16)
Ln(employment-product elasticity)				0.931	0.920	0.916
				(19.49)	(18.97)	(18.96)
Ln(remaining workers)				0.128	0.131	0.130
				(11.17)	(11.39)	(11.32)
[*ln(capital)*]^2				0.012	0.011	0.0112
				(14.82)	(14.39)	(14.20)

[19] As far as is known, only one paper (Escalante, 2004) corrects for endogeneity for the urban Bolivia case, but no one corrects the two biases at the same time.

[*ln(intermediate consumption)*]^2				0.034	0.034	0.034
				(28.72)	(27.69)	(27.63)
D_electric power, gas, and water				0.291	0.275	0.387
				(6.77)	(5.51)	(7.35)
D_construction				0.780	0.780	0.7708
				(18.25)	(18.19)	(18.22)
D_ transport and communication				0.201	0.193	0.196
				(5.44)	(5.16)	(5.27)
D_banking				0.789	0.735	0.7098
				(3.15)	(3.03)	(2.93)

Table 3 continuation

Explanatory variables	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV (2SLS)	IV plus Heckman	OLS	IV (2SLS)	IV plus Heckman
D_business services				0.473	0.442	0.439
				(6.92)	(6.32)	(6.30)
D_ Other social services				0.155	0.171	0.169
				(3.21)	(3.54)	(3.53)
D_has RUC				0.204	0.181	0.1753
				(5.90)	(5.12)	(5.0)
D_credit				0.043	0.041	0.035
				(1.99)	(1.85)	(1.61)
Inverse Mills ratio			-0.374			-0.2283
			(-7.17)			(-5.99)
Constant	4.871	4.372	5.124	4.910	4.657	5.115
	(69.590)	(36.30)	(32.17)	(86.57)	(48.80)	(42.57)
R^2	0.34	0.32	0.33	0.64	0.63	0.64
Observations	4251	4251	4251	4251	4251	4251

Notes: 1) In brackets are the *t*-statistic values calculated from the robust variance-covariance matrix; 2) regressions were estimated from the Household Survey 2003-2004, using monthly data; 3) only the variables significant at the 10% level were included; 4) the *probit* model (for calculating the inverse Mills ratio) included the following significant variables: *D_male*, *D_indigenous*, dummy for household head, dummy for married, dummy for student, and the ratio of children per adult.

Regression (4) presents the estimation using OLS, and (5) and (6) correct, respectively, by endogeneity bias and both biases. In general, all coefficients show the expected signs. In particular, from the last regression the following aspects related to firm characteristics can be highlighted:

- *Ln(remaining workers)* shows a positive and relatively high coefficient, which can be associated with the size of the firm: an increase in 1% in this kind of employment is associated to an earnings increase of approximately 0.13%.

- The *proxy* variable of capital has a positive effect on earnings; however, the elasticity is relatively low on average – $(0.024)–0.011\times2\times$*ln(capital)* [20] – which may be due to both a low use of this factor of production and the *proxy* used that can underestimate the elasticity. It is worth noting, however, that the dispersion of this variable is high (see Tables 1 and 2), leading to a substantially higher elasticity for the highest decile of the sample (0.14).

[20] *Ln(capital)* corresponds to the average sample.

- The earnings-intermediate consumption elasticity shows a high value[21]; with an increase of 1% in this kind of expense is associated to an increase of 0.50% in labor income.

- The credit dummy shows a significant positive coefficient, proving that, ceteris paribus, access to physical and operating capital improves the earnings median by approximately 3.6%, $(e^{0.035}-1)\times 100$.

- The dummy equal to one when the firm is registered in the RUC, *D_has RUC*, has a positive and high impact, increasing the earnings median by approximately 19.2%, $(e^{0.175}-1)\times 100$, reflecting — as pointed out above — a better production base in terms of basic management skills, structures, and processes.

- Lastly, the wage gaps by sectors are significant in some cases, with the estimated high values of the coefficients of banking and construction standing out. These factors would be explaining the differences in market power but also differences in production processes by sector.

The consideration of firm characteristics in the regressions on earnings seems fundamental for independent workers: comparing regressions (3) with (6), two main differences arise. First the coefficient of determination (R^2) almost doubles in regression (6), when firm characteristics are included, even though not all the significant factors of production variables were included because of multi-collinearity problems. Second, the coefficients related with individual characteristics are exposed as highly biased when firm characteristics are excluded, confirming the observation of Abowd *et al.* (1999) and Abowd and Kramarz (1999) regarding to the variables omission problems. In particular, regression (3) predicts that for workers with up to 12 years of schooling, an additional year of schooling would increase earnings by 8.1%, $(e^{0.078}-1)\times 100$, while in (6) the estimated increase reaches only 3.9% (less than half). Education returns for workers with superior education are in both regressions higher compared to the rest of workers, however the gap between the estimations is also high (a difference of 5.2% in the rate of return).

In general, the values of the coefficients of individual characteristics are overestimated when firm characteristics are excluded, which is the result of the relation between the two types of variables (see Abowd *et al.* 1999): intuitively individual characteristics would also explain firm

[21] In this case, elasticity is equal to: $0.034\times 2\times ln(intermediate\ consumption)$, where the variable corresponds to the average sample.

characteristics when these last variables are not included in earnings regressions[22]. For example, independent workers with more years of schooling usually belong to richer families and/or have greater savings from previous jobs, which places them in an advantageous position in the acquisition of factors of production, both in quantity and quality.

Regressions (7) to (12) in Table 4 present the econometric results for dependent workers. The last three regressions include firm characteristics that have been narrowed down by the proxy *ln*(*average wages by sector-size*) and by a dummy for workers subject to labor legislation. In the same way as for independent workers, the Hausman test and the significance of the inverse Mills ratio coefficient show, respectively, that endogeneity and sample selection bias are important, with (8) and (11) being corrected by endogeneity and (9) and (12) by both biases.

Table 4: Ln(Earnings) Determinants per Month for Dependent Workers, 2003-2004

Explanatory variables	(7)	(8)	(9)	(10)	(11)	(12)
	OLS	IV (2SLS)	IV plus Heckman	OLS	IV (2SLS)	IV plus Heckman
Years of schooling up to 12	0.064	0.116	0.096	0.042	0.076	0.059
	(19.30)	(10.02)	(8.23)	(8.83)	(4.33)	(3.20)
Over 12 year of schooling	0.100	0.136	0.128	0.070	0.102	0.096
	(43.99)	(23.19)	(21.70)	(14.04)	(9.98)	(9.67)
Experience	0.055	0.055	0.045	0.043	0.045	0.036
	(23.42)	(22.99)	(17.21)	(11.60)	(10.77)	(9.0)
[*Experience*]^2	-0.0008	-0.0007	-0.0005	-0.0006	-0.0006	-0.0005
	(-15.29)	(-11.90)	(-9.93)	(-9.32)	(-9.03)	(-7.68)
D_male	0.282	0.261	0.127	0.249	0.252	0.130
	(15.12)	(12.97)	(5.02)	(7.64)	(7.95)	(5.25)
D_indigenous	-0.116	-0.009	0.051	-0.114	-0.035	0.022
	(-5.60)	(-0.35)	(1.96)	(-4.31)	(-1.14)	(0.67)
Hours_month	0.002	0.002	0.002	0.002	0.002	0.002
	(16.15)	(16.93)	(16.11)	(6.87)	(6.27)	(5.67)
Ln(average wages by groups)				0.361	0.264	0.248
				(11.36)	(8.04)	(7.57)
D_subject to labor legislation				0.361	0.285	0.279
				(6.15)	(3.95)	(3.71)
Inverse Mills ratio			-0.350			-0.326
			(-8.48)			(-5.73)
Constant	4.771	4.190	4.950	2.498	2.735	3.535

[22] According to Abowd *et al.*, exclusion of firm variables implies that the estimated coefficients of individual variables sum both effects, individual plus (employment weighted average) firm effects, considering that the covariance matrix between these two categories of variables are non-zero.

	(92.850)	(37.780)	(34.22)	(10.16)	(9.04)	(14.48)
R^2	0.44	0.41	0.43	0.50	0.49	0.50
Observations	5608	5608	5608	5608	5608	5608

Notes: 1) In brackets are the *t*-statistic values calculated from the robust variance-covariance matrix; 2) standard errors were corrected by intra-group correlation derived from average wages (46 clusters); 3) regressions were estimated from the Household Survey 2003-2004 using monthly data; 4) only the variables significant at the 10% level were included; 5) the *probit* model (for calculating the inverse Mills ratio) included the following significant variables: *D_male*, *D_indigenous*, dummy for household head, dummy for married, dummy for student, and the ratio of children per adult.

In general, all the coefficients have the expected signs and are significant at least at 10%, with the exception of *D_indigenous*, being in some cases insignificant and in others having the opposite sign of that expected (regression (9)). The inclusion of firms' variables improves the coefficient of determination by more than 13%, and shows that: first — on average — the logarithm of individual wages is composed at a level of 24.8% by the logarithm of earnings by sector and size; and, second, *D_subject to labor legislation* has a positive and high impact, increasing the median of earning by approximately 32.3%, $(e^{0.279}-1)\times 100$, which can be related to both higher individual effort (giving better labor conditions) and better production structures for working.

In the same way as for independent worker, the regressions that only consider individual characteristics present overestimated coefficients. In particular, an increase in one year of schooling is associated to an increase of 10.1% in earnings for workers with up to 12 years of education and by 13.7% for workers with over 12 years in regression (9), with these rates being 6.1% and 10.1% respectively in the last regression (considering firm characteristics). Lastly, in (12) the insignificance of the dummy for "indigenous" may correspond to the inclusion of the variable *ln(average wages by sector-size)*, which in some way controls the segmentation problem by sector where indigenous workers are located in low-paid sectors (see Mercado *et al.* 2003).

As in the case of independent workers, the overestimation of the coefficients of individual characteristics is the result of their relation with firm characteristics. Intuitively, for example, *years of schooling* presents a positive correlation with the size of firms, suggesting that firms are more intensive in skilled workers as they become larger (possibly because they corresponds to sub-sectors more intensive in capital). Additionally, the variable *D_subject to labor legislation* is positively related to *over 12 years of schooling* and negatively related to *years of schooling up to 12*, which suggest that formal firms tend to hire skilled workers because they are both more intensive in this kind of work (such as for example firms belonging to the electricity, natural gas and water supply, public administration, banking, and education sectors), and attract skilled workers even if the job is for unskilled ones.

Finally, Table 5 presents earnings inequality decompositions from the estimated regressions that control for endogeneity and sample selection biases. When the regressions consider only individual characteristics – regressions (3) for independent workers and (9) for dependent ones – the results are similar to the previous findings, where education was the most relevant variable (Fields *et al.* 1998, Andersen 2001, Spatz and Steiner, 2002).

Table 5: Shorrocks and Fields Earnings Inequality Decompositions for Dependent and Independent Workers

Explanatory Variables	Independent Workers				Dependent W		
	Only individual variables		All variables		Only individual variables		
	Shorrocks (a)	Fields (b)	Shorrocks (c)	Fields (d)	Shorrocks (e)	Fields (f)	S
Years of schooling	13.62%	41.16%	6.62%	10.36%	33.16%	76.58%	
Experience	-0.56%	-1.71%	-0.25%	-0.40%	5.06%	11.70%	
D_male	6.34%	19.15%	3.94%	6.16%	0.90%	2.09%	
D_indigenous	1.51%	4.57%	1.17%	1.84%	-0.37%	-0.86%	
Hours_month	12.21%	36.89%	5.96%	9.33%	0.98%	2.26%	
Production variables			40.32%	63.10%			
Ln(average wages by sector-size)							
Dummies of sectors			2.92%	4.57%			
D_subject to labor legislation							
D_has RUC			2.12%	3.31%			
D_credit			0.17%	0.26%			
Inverse Mills ratio	-0.02%	-0.07%	0.93%	1.46%	3.56%	8.22%	
Residuals	66.90%		36.10%		56.70%		
Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	

Notes: 1) The first two columns are calculated from regression (3) from Table 3 and the second two from (6); 2) columns (e) and
regression (9) of Table 4, and the last two columns from (12); 3) *years of schooling* variables were considered instead of their c
variables; and 4) *years of schooling* aggregates the variables *years of schooling up to 12* and *over 12 year of schooling*, *Experience* ag
and in square, and *Production variables* aggregates the *ln(product-labor elasticity)* and the factor of production variables.

In this regard, the Fields decomposition (columns (b) and (f) of Table 5) shows that *years of schooling* explains 41.2% of the explained earnings inequality for independent workers and 76.6% for dependent ones. However, these high percentages are partly explained by the exclusion of the residuals in the Fields calculation, which are relevant in both cases (66.9% and 56.7% respectively). Therefore, using the Shorrocks-Fields decomposition, the variable fails to explain 13.6% of earnings inequality for independent workers and 33.2% for dependent ones.

The calculation of the regressions that also considers firm characteristics – (6) and (12) – exposes a different scenario, mainly for independent workers. In this case, the Shorrocks decomposition (column (c) of Table 5) shows that *production variables – Ln(product-labor elasticity)* plus the factor of production variables – explain 40.3% of the earnings dispersion, and all firm characteristics, where together *production variables*, *Dummies of sectors*, *D_has RUC* and *D_credit*, explain 45.5% (71.2% considering the Fields decomposition). Individual characteristics lose their explicative power; in particular, the coefficient of *years of schooling* falls to about half, explaining now only 6.6% of earnings inequality.

Lastly, in the case of dependent workers, firm characteristics *proxies* jointly explain 11.8% of earnings. *Years of schooling* also loses its explicative power from 33.16% to 28.35% (columns (e) and (g)); however, it continues to be the most important factor, which can be attributed to both non-detailed available information at the firm level for this population and to the relatively higher importance of education for these workers.

In short, the decompositions show that firm characteristics for explaining earnings inequality are relevant, being essential for independent workers. Additionally, education loses its explicative power when firm characteristics are included in the regressions, which can correspond – as discussed above – to its relation with firm characteristics.

V. Conclusions

This paper analyzes the factors that explain earnings in the urban areas of Bolivia, considering both individual variables (usually related to the human capital theory) and firm characteristics. Initially it can be observed on average — through simple charts — that there is a positive relationship between education and earnings under a non-linear structure; however, highlighting a high dispersion. Two variables at the firm level, sector and size, are then added to try to explain part of these dispersions, which are not accurately derived by differences (or similarities) in years of schooling.

In this regard, a new simple methodology is proposed, adapted to the information presented in the household surveys, to include firm characteristics in earnings regressions. This methodology

relaxes the hypothesis of a perfect labor market and merges a simple specification of the human capital theory and the firms' optimization problem from which follows the labor marginal productivity value. Two simple models are derived from this methodology to be analyzed empirically. The first one corresponds to independent workers, representing 43.1% of the total workers studied, for which there is information on production costs and other firm characteristics (besides the usual individual variables), allowing a detailed evaluation of earnings determinants. The second one is for dependent workers, where firm characteristics are approximated by sector and size as well as by the legal condition of the workers.

Earnings are estimated in levels and variability considering the decomposition analysis of inequality of Shorrocks (1982) and Fields (2002). The empirical analysis confronted three econometric problems: multi-collinearity, endogeneity and sample selection bias. Multi-collinearity arises for the model of independent workers because firm characteristics were derived from a translog production function, which has various variables related with the factors of production. In this regard, the problem was corrected by using the Bierrens (2007) procedure, by evaluating the robustness of the explanatory variables, and finally by including only the factors of production variables that have a low correlation between them and have the highest t-values (in order to not overestimate the firm characteristics or the coefficient of determination used for the inequality decomposition analysis). The endogeneity problem arises because, according to the human capital theory, workers' ability is an omitted variable that is correlated to years of schooling. Consequently, instrumental variables were used, estimating the models by Two-Stage Least Squares (2SLS). Lastly, the sample selection bias arises because the division of the employed population between independent workers and dependent ones implies the use of non-randomly selected samples leading to missing data problems (see Heckman, 1979). To correct this bias simultaneously with the endogeneity problem, the Wooldridge (2002) approach was used, which combines Heckman's two-step procedure and instrumental variables.

The empirical analysis leads to the results that follow for independent workers. First, firm characteristics are exposed as fundamental factors for explaining earnings, highlighting the intermediate consumption and (remaining) employment, where an increase of 1% is related to an increase of 0.50% and 0.13% respectively. Additionally, the dummies for credit and firm's registration in the RUC increase the median of earnings by 3.6% and 19.2% respectively, reflecting the importance of access to physical and operating capital, as well as better production bases (management skills, structures, and processes).

Second, firm characteristics seem fundamental for explaining earnings in the urban Bolivia case: the coefficient of determination (R^2) almost doubles when these variables are included in the regressions. Additionally, the estimated coefficients related to individual characteristics are

exposed as highly biased (overestimated) when only these variables are considered, confirming the observation of Abowd *et al.* (1999) and Abowd and Kramarz (1999) regarding the variables omission problems. In particular, the education rate of return is estimated at 8.1% for workers with up to 12 years of schooling and 9.4% for the rest (over 12 year of schooling) compared to the estimated rates of 3.9% and 4.2% when firm characteristics are included. Finally, the Shorrocks-Fields decomposition shows that individual factors lose their explicative power when firm characteristics are considered. Additionally, these last variables jointly explain 45.5% of earnings inequality (71.2% considering the Fields decomposition), followed by education, that explains 6.6%.

In the case of dependent workers, the results that follow arise. First, the inclusion of firms' variables improves the coefficient of determination by more than 13%, and shows that — on average and in logarithms — 24.8% of wages are composed by earnings by sector and size, and the median of earnings increases to approximately 32.3% for workers subject to labor legislation. Second, in the same way as for independent workers, the regressions that only consider individual factors present overestimated coefficients. In particular, an increase in one year of schooling is associated to an increase in 10.1% of earnings for workers with up to 12 years of education and in 13.7% for workers with over 12 years, and decreases to 6.1% and 10.1% when firm characteristics are included. In this case individual characteristics would also represent firm characteristics in earnings regressions where the last variables are not considered.

Lastly, the Shorrocks-Fields decomposition analysis shows that years of schooling remains in first place, explaining 28.35% of earnings inequality, followed by firm characteristics proxies that jointly explain 11.8%, which can be attributed to both non-detailed available information at the firm level for this population, and to the relatively higher importance of education for these workers.

In short, the empirical analysis developed shows that firm characteristics, as determinants of earnings in levels and in inequality, are fundamental for independent workers and relevant for dependent workers; and that for both categories of workers their exclusion leads to highly biased estimations in earnings regressions. These new findings represent a new contribution to the empirical literature on earnings determinants for urban Bolivia, as well as to the vision regarding low wages, inequality and poverty problems.

References

Abowd, J. M. and F. Kramarz (1997), "Internal and External Labor Markets: An Analysis of Matched Longitudinal Employer-Employee Data", National Bureau of Economic Research, NBER Working Papers 6109.

Abowd, J. M. and F. Kramarz (1999), "The Analysis of Labor Markets using Matched Employer-Employee Data", *Handbook of Labor Economics*, O. Ashenfelter and D. Card (eds.), chapter 40, volume 3B, (Amsterdam, North Holland): 2629-2710.

Abowd, J. M., F. Kramarz and D. Margolis (1999), "High Wage Workers and High Wages Firms", *Econometrica* 67(2): 251 – 333.

Abowd, J. M., F. Kramarz, D. and S. Woodcock (2008), "Econometric Analyses of Linked Employer-Employee Data," *The Econometrics of Panel Data*, L. Matyas and P. Sevestre (eds.), pp 727-760.

Andersen, L. E. (2001), "Wage Differentials between Bolivian Cities", *Economía Política* II(1): 48-62.

Andersen, L. E. and B. Muriel H. (2002), "Cantidad versus Calidad en Educación: Implicaciones para Pobreza", Instituto Nacional de Estadística, *Revista de Estudios Económicos y Sociales* (1): 8-41.

Aydemir, A. and M. Skuterud (2008), "The Inmigrant Wage Differential within and across Establishments", *Industrial and Labor Relations Review* 61(3): 333-352.

Bayard, K. and K. Troske (1999), "Examining the Employer-Size Wage Premium in the Manufacturing, Retail Trade, and Services Industries using Employer-Employee Matched Data", *American Economic Review* 89(2): 99-103.

Becker G. (1964), *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, Chicago, University of Chicago Press (1993-3rd ed.).

Berndt, E. R. and L. R. Christensen (1973), "The Translog Function and the Substitution of Equipment, Structures, and Labor in U.S. Manufacturing 1929–68", *Journal of Econometrics* 1: 81–114.

Bierens, H. J. (2007), "Multicollinearity", Pennsylvania State University, mimeo.

Binswanger, H. P (1974), "The Measurement of Technical Change Biases with Many Factors of Production", *American Economic Review* 64(6): 963-976.

Breau, S. and D. L. Rigby (2006), "Is There Really an Export Wage Premium? A Case Study of Los Angeles Using Matched Employee – Employer Data", *International Regional Science Review* 29(3): 297-310.

Brown, C. and J. L. Medoff (1989), "The Employer Size-Wage Effect", National Bureau of Economic Research, NBER Working Paper 3336.

Card, D. (1999), "The Causal Effect of Education on Earnings", *Handbook of Labor Economic*, O. Ashenfelter and D. Card (eds.), chapter 30, volume 3A, (Amsterdam, North Holland): 1801-1863.

Chen G. and H. Shigeyuki (2009), "Economic Returns to Schooling in Urban China: OLS and the Instrumental Variables Approach", *China Economic Review* 20: 143–152.

Cowell, F. and C. Fiorio (2006), "Rethinking Inequality Decomposition: Comment", London School of Economics, mimeo.

Cowell, F. and C. Fiorio (2009), "Inequality Decompositions: A Reconciliation", Society for the Study of Economic Inequality, Working Paper 117.

Das, M., W.K. Newey and F. Vella (2003). "Nonparametric Estimation of Sample Selection Models", *Review of Economic Studies* 70: 33-58.

El–Attar, M. and E. Lopez–Bazo (2008), "Firm Size and the Entire Distribution of Wages: A Semiparametric Approach", *Applied Economics Letter* 15: 427-430.

Escalante, C. E. (2004), "Los Retornos de la Inversión en Capital Humano en Bolivia", Unidad de Análisis de Políticas Económicas y Sociales, *Revista de Análisis Económico* 19: 1-26.

Fields, G. S. (2002), "Accounting for Income Inequality and its Change: A New Method, with Application to the Distribution of Earnings in the United States", Cornell University, Ithaca, NY, Articles and Chapters, paper 265.

Fields, G. S., J. B. Leary, L. F. Lopez-Calva and E. Perez-de-Rada (1998), "Education's Crucial Role in Explaining Labor Income Inequality in Urban Bolivia", Development Discussion Paper No. 658, Harvard Institute for International Development, Octover.

Garcia, J., P.J. Hernandez, and A. Lopez-Nicolas (2001), "How Wide is the Gap: An Investigation of Gender Wage Differentials using Quantile Regressions", *Empirical Economics* 26(1): 149-169.

Haltiwanger, J. C., J. I. Lane and J. R. Spletzer (2000), "Wages, Productivity, and the Dynamic Interaction of Businesses and Workers", National Bureau of Economic Research, NBER Working Paper No 7994.

Hamermesh, D. (2008), “Fun with Matched Firm – Employee Data: Progress and Road Maps”, *Labour Economics* 15: 663–673.

Hassink, W. and G. Russo (2008), “Wage Differences between Internal and External Candidates”, *International Journal of Manpower* 29(8): 715-730.

Heckman, J. J. (1979), “Sample Selection Bias as a Specification Error”, *Econometrica* 47(1): 153-161.

Hellerstein, J., D. Neumark and K. Troske (1999), “Wages, Productivity, and Worker Characteristics: Evidence from Plant – Level Production Functions and Wage Equations”, *Journal of Labor Economics* 17(3): 409 - 446.

Hellerstein, J. and D. Neumark (2004), “Production Function and Wage Equation Estimation with Heterogeneus Labor: Evidence from a new Matched Employer – Employee Data Set”, National Bureau of Economic Research, NBER Working Paper No 10325.

Herrera, J. (2003), “Análisis de la Pobreza en el Perú 2002 desde la Perspectiva de los Hogares y las Unidades de Producción” (resultados preliminares), IRD – CIPRE, Octover.

Ilmakunnas, P. and M. Maliranta (2005), “Technology, Labour Characteristics and Wage–productivity Gaps”, *Oxford Bulletin of Economics and Statistics* 67(5): 623-645.

Israeli, O. (2007), “A Shapley-based Decomposition of the R-Square of a Linear Regression, *Journal of Economic Inequality* 5:199–212.

Jones, P. (2001), “Are educated Workers Really more Productive?”, *Journal of Development Economics*, 64:57-79.

Lalleman, T., R. Plasman and F. Rycx (2005), “Why do Large Firms Pay Higher Wages? Evidence from Matched Worker–firm Data”, *International Journal of Manpower* 26 (7/8): 705 – 723.

Martins, P. (2008), “Dispersion in Wage Premiums and Firm Performance”, Centre for Globalization Research, Working Papers No. 8.

Menezes-Filho, N., M. A. Muendler and G. Ramey (2006), “The Structure of Worker Compensation in Brazil, With a Comparison to France and The United States”, mimeo.

Mercado, A., L. E. Andersen and B. Muriel H. (2003), “Discriminación Étnica en el Sistema Educativo y en el Mercado de Trabajo de Bolivia”, *Revista Latinoamericana de Desarrollo Económico* 1(sep): 69-98.

Mincer, J. (1974), *Schooling, Experience and Earnings*, Nueva York: NBER Press.

Moensted, T. (2000), "Wage Differentials between the Formal and the Informal Sector in Bolivia", IISEC-UCB Working Paper 2/2000, La Paz-Bolivia.

Morduch, J. and T. Sicular (2002), "Rethinking Inequality Decomposition, with Evidence from Rural China", *The Economic Journal* 112 (January): 93-106.

Moretti, E. (2004), "Workers' Education, Spillovers, and Productivity: Evidence from Plant – Level Production Functions", *American Economic Review* 94(3): 656-690.

Muravyev, A. (2009), "Employer size, Wages and Unobserved Skills: Evidence from Moonlighters in the UK", *The Manchesteer School* 77(6): 651–674.

Muriel, H. B. (2005a), "Female Labor Market Conditions in Urban Bolivia", Maestrías para el Desarrollo, Universidad Católica Boliviana "San Pablo", Working Paper 003/2005, La Paz, Bolivia.

..... (2005b), "Female Labor Force Participation in the Urban Bolivia", Maestrías para el Desarrollo, Universidad Católica Boliviana "San Pablo", Working Paper 001/2005, La Paz, Bolivia.

..... (2010), "Eficiencia versus Equidad en la Producción Industrial", Newspaper Nueva Economía (4 de Abril), La Paz, Bolivia.

Muriel, H. B. and L.C. Jemio (2010), "Mercado Laboral y Reformas en Bolivia", *El Mercado Laboral en Bolivia*, B. Muriel H. (eds.), Universidad Católica Boliviana "San Pablo" (UCB) - International Development Research Centre (IDRC), La Paz, Bolivia.

Oi, W.Y. and T.L. Idson (1999), "Firm-Size and Wages", *Handbook of Labor Economics*, O. Ashenfelter and D. Card (eds.), chapter 33, volume 3B, (Amsterdam, North Holland): Pages 2165-2214.

Perez de Rada, E. (1997) "Discriminación Salarial por Género y Etnia en Ciudades Principales de Bolivia", UDAPSO Working Paper 47/97, La Paz-Bolivia.

Pylkkänen, E. (2000), "Modeling Wages and Hours of Work", Ministry of Finance, Stockholm.

Ramírez, R. (2003), "Pobreza y Discriminación Laboral por Género", Bachelor Thesis in Economics, Universidad Católica Boliviana, La Paz-Bolivia.

Reza A. G. and A. Mussurov (2006), "Returns to Schooling in Kazakhstan: OLS and Instrumental Variables Approach", The Institute for the Study of Labor, IZA Discussion Paper No. 2462 (November).

Rivero, R. and W. Jiménez (1999) "Diferencias Salariales en el Mercado de Trabajo Urbano en Bolivia, 1981-1997", Unidad de Análisis de Políticas Económicas y Sociales, *Revista de Análisis Económico* 17: 15-45.

Soderbom, M., T. Francis and A. Wambugu (2005), "Unobserved Heterogeneity and the Relation Between Earnings and Firm Size: Evidence from two Developing Countries", *Economics Letters* 87:153 – 159.

Shorrocks, A. F. (1982), "Inequality Decomposition by Factor Components", *Econometrica* 50(1): 193-211.

Spatz, J. and S. Steiner (2002), "Post-Reform Trends in Wage Inequality: The Case of Urban Bolivia", Kiel Working Paper 1126.

Temesgen, T. (2005), "Determinants of Wage Structure and Returns to Education in a Developing Country: Evidence from Linked Employer – Employee Manufacturing Survey Data of Ethiopia", *Seoul Journal of Economics* 18(4): 277-203.

Trefler, D. (1993), "International Factor Price Differences: Leontief was Right", *Journal of Political Economy* 101(6): 961-87.

Troske, K. R., (1995), "The Worker – Establishment Characteristics Database", Center for Economic Studies, U.S. Census Bureau.

Verner, D. (s.f.) "Wage and Productivity Gaps: Evidence from Ghana", mimeo.

Verner, D. (1999), "Are Wages and Productivity in Zimbabwe Affected by Human Capital Investment and International Trade?", World Bank, Policy Research Working Paper No 2101.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA MIT Press.

Woodcock, S. (2007), "Match Effect", MPRA Paper 154.

Woodcock, S. (2008), "Wage Differentials in the Presence of Unobserved Worker, Firm and Match Heterogeneity", *Labour Economics* 15: 772-794.

Yasar, M. and C. Morrison (2008), “Capital–skill Complementarity, Productivity and Wages: Evidence from Plant–level Data for a Developing Country”, *Labour Economics* 15:1 – 17.

Annex

Table A.1: Ln(Earnings) Determinants per Month for Independent Workers, 2007

Explanatory variables	(13)	(14)	(15)	(16)	(17)	(18)
	OLS	IV (2SLS)	IV+Heck.	OLS	IV (2SLS)	IV+Heck.
Years of schooling	0.047	0.080	0.081	0.017	0.029	0.030
	(7.28)	(8.29)	(8.46)	(3.45)	(3.60)	(3.74)
Experience	0.019	0.022	0.016	0.014	0.015	0.011
	(3.05)	(3.62)	(2.37)	(3.08)	(3.37)	(2.38)
[*Experience*]^2	-0.0002	-0.0002	-0.0001	-0.0002	-0.0002	-0.0001
	(-2.24)	(-1.82)	(-0.97)	(-2.43)	(-2.26)	(-1.52)
D_male	0.4679	0.4169	0.3821	0.2348	0.2193	0.1971
	(9.81)	(8.49)	(7.41)	(5.78)	(5.30)	(4.52)
D_indigenous	-0.137	-0.056	-0.087	-0.130	-0.105	-0.1236
	(-2.56)	(-0.97)	(-1.45)	(-3.46)	(-2.62)	(-2.93)
Hours_month	0.003	0.003	0.003	0.001	0.001	0.001
	(10.53)	(10.26)	(10.07)	(5.27)	(5.34)	(5.24)
Ln(product-labor elasticity)				0.910	0.908	0.908
				(22.53)	(22.60)	(22.53)
Ln(remaining workers)				0.521	0.520	0.5223
				(8.83)	(8.92)	(8.95)
Ln(remaining workers)				-0.039	-0.039	-0.040
×*ln(intermediate consumption)*				(-3.63)	(-3.65)	(-3.68)
[*ln(capital)*]^2				0.014	0.014	0.0137
				(11.42)	(10.72)	(10.59)
[*ln(intermediate consumption)*]^2				0.033	0.033	0.033
				(26.50)	(26.34)	(26.28)
D_construction				0.866	0.869	0.870
				(10.96)	(11.01)	(11.11)
D_ transport and communication				0.238	0.234	0.2351
				(4.53)	(4.45)	(4.47)
D_banking				2.589	2.590	2.614
				(40.15)	(40.25)	(38.93)
D_business services				0.568	0.517	0.5190
				(6.16)	(5.37)	(5.38)
Inverse Mills ratio			-0.202			-0.122
			(5.18)			(-1.68)
Constant	5.229	4.820	5.179	5.085	4.952	5.166
	(43.79)	(33.43)	(22.13)	(55.59)	(44.04)	(30.37)
R^2	0.22	0.20	0.20	0.60	0.60	0.60

Observations	1615	1615	1615	1615	1615	1615

Notes: 1) in brackets are the *t*-statistic values calculated from the robust variance-covariance matrix; 2) the years of schooling showed linear tendency for this group of workers, even though the two instrumental variables were used; 3) the *probit* model (for calculating the inverse Mills ratios) included the following significant variables: *D_male*, *D_indigenous*, dummy for household head, dummy for married and a dummy for student.

Table A.2: Ln(Earnings) Determinants per Month for Dependent Workers, 2007

Explanatory variables	(19)	(20)	(21)	(22)	(23)	(24)
	OLS	IV (2SLS)	IV+Heck.	OLS	IV (2SLS)	IV+Heck.
Years of schooling up to 12	0.056	0.104	0.078	0.036	0.069	0.047
	(10.96)	(5.18)	(3.84)	(3.77)	(2.97)	(1.95)
Over 12 years of schooling	0.080	0.112	0.101	0.054	0.084	0.076
	(23.26)	(10.83)	(9.83)	(7.37)	(6.35)	(5.92)
Experience	0.051	0.051	0.040	0.042	0.044	0.034
	(16.26)	(16.15)	(11.84)	(12.23)	(12.63)	(10.95)
[*Experience*]^2	-0.0007	-0.0006	-0.0005	-0.0006	-0.0005	-0.0004
	(-10.25)	(-7.34)	(-5.88)	(-8.14)	(-6.48)	(-4.90)
D_male	0.322	0.300	0.149	0.261	0.262	0.126
	(12.03)	(10.11)	(4.16)	(7.63)	(7.31)	(2.74)
D_indigenous	-0.136	-0.070	0.033	-0.139	-0.083	0.011
	(-3.86)	(-1.74)	(0.81)	(-4.36)	(-2.55)	(0.29)
Hours_month	0.002	0.002	0.002	0.002	0.002	0.002
	(9.94)	(10.64)	(10.38)	(5.24)	(5.33)	(4.89)
Ln(average wages by groups)				0.371	0.298	0.284
				(5.93)	(3.90)	(4.02)
D_subject to labor legislation				0.283	0.207	0.190
				(4.69)	(3.19)	(2.94)
Inverse Mills ratio			-0.419			-0.377
			(-7.80)			(-5.79)
Constant	5.251	4.713	5.596	2.797	2.898	3.777
	(69.80)	(24.380)	(24.82)	(5.25)	(4.11)	(6.35)
R^2	0.39	0.37	0.40	0.46	0.44	0.46
Observations	2433	2433	2433	2433	2433	2433

Notes: 1) in brackets are the *t*-statistic values calculated from the robust variance-covariance matrix; 2) the *probit* model (for calculating the inverse Mills ratios) included the following significant variables: *D_male*, *D_indigenous*, dummy for household head, dummy for married, a dummy for student, a dummy for migrant and the ratio of children per adult.